(3)

EXCLUSIVE CONTROL SYSTEM FOR DISTRIBUTED DATABASE

Patent Number: JP10040154
Publication date: 1998-02-13
Inventor(s): HIRAMATSU TATSUO
Applicant(s):: MEIDENSHA CORP
Requested Patent: JP10040154
Application Number: JP19960195804 19960725
Priority Number(s):
IPC Classification: G06F12/00 ; G06F12/00
EC Classification:
Equivalents:

Abstract

PROBLEM TO BE SOLVED: To provide an exclusive control system for a distributed database which increase the reference speed of data and facilitates data management without increasing a communication load and loads on other computers.
SOLUTION: A key consisting of an identifier, an exclusive level, etc., for data is given to respective computers (x), (y), and (z), and this key allows only a computer which has the key to operate data, so that the respective computers are allowed to operate the data by transferring, lending, collecting, and returning the key. Computers are given a key which allows only a read to enable parallel reference, information for authentication is added to the key to limit the transfer and lending, and a matrix of transfer or lending is added to the key to make it possible to reserve operation; and the exclusive level is increased on collecting the key.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（A）　(11)特許出願公開番号

特開平10－40154

(43)公開日　平成10年（1998）2月13日

(51)Int.Cl.⁶	識別記号	庁内整理番号	ＦＩ		技術表示箇所
Ｇ０６Ｆ　12/00	５３５		Ｇ０６Ｆ 12/00	５３５	Ｚ
	５４５			５４５	Ａ

審査請求　未請求　請求項の数１　ＯＬ（全５頁）

(21)出願番号　特願平8－195804

(22)出願日　平成8年（1996）7月25日

(71)出願人　000006105
株式会社明電舎
東京都品川区大崎2丁目1番17号

(72)発明者　平松　辰夫
東京都品川区大崎2丁目1番17号　株式会社明電舎内

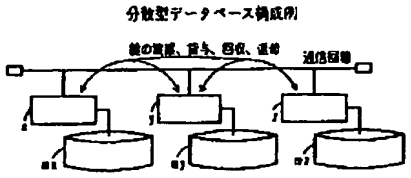
(74)代理人　弁理士　志賀　富士弥　（外１名）

(54)【発明の名称】分散型データベースの排他制御方式

(57)【要約】

【課題】　通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおいて、サーバやロックサーバによる排他制御では処理効率が悪くなる。

【解決手段】　データの識別子と排他レベルなどからなる鍵を各計算機ｘ，ｙ，ｚに付与し、この鍵は、鍵を所持している計算機だけにデータ操作を許可し、鍵の譲渡、貸与、回収、返却で各計算機によるデータ操作を可能にする。読み出しのみが可能な鍵を複数の計算機に貸与することで並列参照を可能とし、鍵に認証のための情報を付加することで、譲渡や貸与に制限を加え、鍵に譲渡あるいは貸与の待ち行列を付加することで、操作の予約を可能とし、鍵を回収することによる排他レベルの格上げを可能とする。



1

【特許請求の範囲】

【請求項1】　通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおいて、

データベース中の各データの操作について、データの識別子と排他レベルなどからなる鍵を各計算機に付与し、前記鍵は、以下の使用方法、

・必要な排他レベルを満足する鍵を所持している計算機だけにデータ操作を許可する。

・鍵を所持していない計算機は、マルチキャストなどの通信方式を用いて鍵の譲渡あるいは貸与を要求する。

・鍵を所持している計算機は、それを譲渡あるいは貸与することができる。

・読み出しのみが可能な鍵を複数の計算機に貸与することで並列参照を可能とする。

・鍵を貸与された計算機はトランザクションの終了時などに鍵を返却する。

・書き込み可能な鍵を移譲することにより、データに対する優先操作権を放棄することを可能とする。

・鍵に認証のための情報を付加することで、譲渡や貸与に制限を加えることを可能とする。

・鍵に譲渡あるいは貸与の待ち行列を付加することで、操作の予約を可能とする。

・複写のための読み出し鍵を用いることで、複製データベースの生成を可能とする。

・鍵を回収することによる排他レベルの格上げを可能とする。

・データの更新値をマルチキャストなどの通信方式を用いて配布する。

・マルチキャストなどの通信方式を用いてデータの更新値を受信することにより複製データベース管理者となり得る。とすることを特徴とする分散型データベースの排他制御方式。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、データベース管理システムにおけるデータ操作の排他制御方式に係り、特に通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおけるデータの更新・参照の衝突でデータの破壊と矛盾が発生するのを防止する分散型データベースの排他制御方式に関する。

【0002】

【従来の技術】分散型データベースには、以下の方式がある。

【0003】（1）サーバ／クライアント方式。

【0004】・データベース中の各データについてサーバとなる計算機を用意する。

【0005】・クライアントとなる計算機は各データについて、サーバである計算機の所在を知ることができる。

【0006】・クライアントである計算機はサーバである計算機に要求することでデータの転送を行い、必要なデータを入手する。

2

【0007】・サーバである計算機によりデータ操作に関する排他制御が行われる。

【0008】（2）ロックサーバ方式。

【0009】・データベース中の各データについて各計算機が直接操作する手段を持つ。

【0010】・データ操作に関する排他制御を行う専用のロックサーバを用意する。

【0011】・各計算機はロックサーバである計算機の所在を知ることができる。

【0012】・各計算機はロックサーバの許可を得た後に、直接データ操作を行う。

【0013】・他の計算機上にあるデータベース中のデータを直接操作する手段としては、NFSによるリモートマウントなどの方式が用いられる。

【0014】（3）その他の方式。

【0015】・排他制御は行わず、操作に付された時刻印を比較することで操作の時系列上の矛盾を検出し、過去の特定の操作をキャンセルすることによって対処する多版時刻印方式がある。

【0016】・特定用途向けに、代表者書き込み、最新書き込み有効、などの性質を利用した方式がある。

【0017】

【発明が解決しようとする課題】従来の（1）の方式では、データベース中の各データについてサーバとなる計算機との対応表を管理し、データ参照要求の前にデータの所在を参照する必要がある。多数の計算機による分散データベースではデータの所在を管理する部分に負荷が集中し、処理の効率を低下させる要因となり得る。

【0018】従来の（2）の方式では、データ操作に関する排他制御の負荷が特定のロックサーバに集中し、処理の効率を低下させる要因となり得る。

【0019】また、従来の（1）、2）の方式共に、基本的にはデータ操作時に必要なデータを取得する方式であるため、データベースが他の計算機上にあるとき、書き込みより読み出しの多い応用においては通信回線を用いたデータ転送が頻繁に起こり、処理の性能を低下させる。また、通信回線の負荷がボトルネックとなり得る。

【0020】従来のその他の方式のうち、多版時刻印方式は多くの版を管理するために多くの主記憶あるいは2次記憶を要求し、また不要になった版を回収するためのガベージコレクションの負荷が大きい。

【0021】代表者書き込み、最新書き込み有効、などの性質は特定の用途以外には利用できないなどの問題がある。

【0022】以上のように幾つかの排他制御の方式があるが、それぞれ問題と思われる点があり、特に多数の計算機が互いに同一のデータを何回も参照したり、時系列的にデータが変化していくようなデータベースの応用に

は向かない面がある。

【0023】本発明の目的は、通信負荷や他の計算機の負荷を上げることなくデータの参照速度を高め、データ管理も容易にする分散型データベースの排他制御方式を提供することにある。

【0024】

【課題を解決するための手段】本発明は、通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおいて、特に多数の計算機が互いに同一のデータを何回も参照したり、時系列的にデータが変化していくようなデータベース応用の場合に効率を良くするデータ操作の排他制御方式とするため、通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおいて、データベース中の各データの操作について、データの識別子と排他レベルなどからなる鍵を各計算機に付与し、前記鍵は、以下の使用方法、

・必要な排他レベルを満足する鍵を所持している計算機だけにデータ操作を許可する。

【0025】・鍵を所持していない計算機は、マルチキャストなどの通信方式を用いて鍵の譲渡あるいは貸与を要求する。

【0026】・鍵を所持している計算機は、それを譲渡あるいは貸与することができる。

【0027】・読み出しのみが可能な鍵を複数の計算機に貸与することで並列参照を可能とする。

【0028】・鍵を貸与された計算機はトランザクションの終了時などに鍵を返却する。

【0029】・書き込み可能な鍵を譲渡することにより、データに対する優先操作権を放棄することを可能とする。

【0030】・鍵に認証のための情報を付加することで、譲渡や貸与に制限を加えることを可能とする。

【0031】・鍵に譲渡あるいは貸与の待ち行列を付加することで、操作の予約を可能とする。

【0032】・複写のための読み出し鍵を用いることで、複製データベースの生成を可能とする。

【0033】・鍵を回収することによる排他レベルの格上げを可能とする。

【0034】・データの更新値をマルチキャストなどの通信方式を用いて配布する。

【0035】・マルチキャストなどの通信方式を用いてデータの更新値を受信することにより複製データベース管理者となり得る。

【0036】とすることを特徴とする。

【0037】

【発明の実施の形態】図1は、分散型データベースを示し、複数の計算機x，y，zがそれぞれ通信回線で結合され、各計算機x，y，zにはデータベースとしてデータを分散して記憶するための記憶装置mx，my，mzを持つ。

【0038】この構成において、各計算機x，y，zは、データベース中の各データについて、データの識別子と排他レベルなどからなる鍵を所持できるようにする。各計算機x，y，zが所持する鍵は、データベースに対してデータの読み出し可能なものや書込み可能なものなどの機能を持たせ、さらに鍵の譲渡・貸与・回収・返却などの機能を持たせ、以下の使用方法でデータベースに対する排他制御がなされる。

【0039】・必要な排他レベルを満足する鍵を所持している計算機だけが、そのレベルに応じたデータ操作を行うことができる。

【0040】・必要な排他レベルを満足する鍵を所持していない計算機は、マルチキャストなどの通信方式を用いて不特定の計算機に対し鍵の譲渡あるいは貸与を要求する。

【0041】・鍵を所持している計算機は、それを譲渡あるいは貸与することができる。

【0042】・読み出しのみが可能な鍵を複数の計算機に貸与することで、トランザクションの並列性を確保することができる。貸与された計算機はトランザクションの終了時などに鍵を返却しなければならない。

【0043】・書き込み可能な鍵を譲渡することにより、対応するデータに対する優先的な操作権を完全に放棄することができる。

【0044】・鍵に認証のための情報を付加することで、譲渡や貸与に制限を加えることができる。

【0045】・鍵に譲渡あるいは貸与の待ち行列を付加することで、操作を予約することが可能となる。

【0046】・プログラム中の任意の時点で複写のための読み出し鍵を譲渡することにより、データベースの複製を生成することが可能となる。

【0047】・鍵を回収する機能を設けることで、排他レベルの格上げを実現することができる。

【0048】・トランザクションの完了時にデータの更新値をマルチキャストなどの通信方式を用いて配布することによりデータの更新値が通信回線上を1度だけ流れるだけで済み、また参照時にデータ転送を行わなくても良い。

【0049】・トランザクションの完了時にデータの更新値をマルチキャストなどの通信方式を用いて配布することにより、各計算機が任意の時点から他の計算機の負荷を増すことなく、複製データベース管理者となり得る。

【0050】このような機能・性質を持つ鍵を使った分散型データベースの使用例を以下に説明する。

【0051】図2の（a）には、計算機xがデータの型Aを定義し、マルチキャストなどの通信方式により他の計算機y，zに型Aを配布する様子を示す。この型Aの定義情報の配布により、記憶装置mx，my，mzにはそれぞれ型Aのデータを格納する領域が確保される。

5　　6

【0052】図2の（b）には、計算機xが型Aのデータa_iを作成・配布する様子を示し、各記憶装置mx，my，mzには型Aのデータa_iがマルチキャストなどの通信方式によりそれぞれ格納される。このとき、データa_iに対する読み出し・書込み等を管理するための鍵は配布元の計算機xが所持する。

【0053】図3は、読み出し鍵を使用して計算機yがデータa_iの読み出しを行う手順を示す。同図の（a）では、計算機yが計算機xに対してデータa_iの読み出し鍵を要求する。この要求に対して、（b）では、計算機xが計算機yにデータa_iの読み出し鍵を貸与する。（c）では鍵を貸与された計算機yがデータa_iを読み出し、そのデータa_iを参照する。（d）では読み出しを終了した計算機yがトランザクションの終了時など適当な時期にデータa_iの読み出し鍵を計算機xに返却する。

【0054】したがって、データa_iの参照は、鍵を所持する計算機xの管理の元に行われ、他の計算機y，zは計算機xに対して読み出し鍵が貸与されることで可能となる。このとき、複数の計算機に対して読み出し鍵を貸与でき、複数の計算機による読み出しが可能となる。貸与された鍵は、データ参照後に返却される。また、貸与される鍵に待ち行列を付加することでデータ参照に優先順位を与えることや、認証のための情報付加で貸与に制限を加えることができる。

【0055】図4は、書込み鍵を使用して計算機zがデータa_iに対する書込みを行う手順を示す。同図の（a）では、計算機zが計算機xに対してデータa_iの書込み鍵を要求する。この要求に対して（b）では計算機xが計算機zにデータa_iの書込み鍵を譲渡する。この譲渡では計算機xにはデータa_iの書込み鍵を持たない状態となり、データa_iに対する優先点きな操作権を放棄する。（c）では鍵を譲渡された計算機zがデータa_iに対して書込みをし、この書込みで更新したデータa_i'を他の計算機x，yに配布し、それぞれの計算機の記憶装置mx，my，mzのデータa_iが更新される。

【0056】したがって、データa_iの更新は、書込み鍵を持つ計算機によりなされ、更新されたデータが各計算機のデータベースに反映される。この書込み鍵の譲渡にも待ち行列の付加や認証情報の付加を行うことができる。

【0057】

【発明の効果】以上のとおり、本発明によれば、通信回線により結ばれた計算機ネットワーク上で動作する分散型データベースにおいて、読み出し鍵や書込み鍵などの各種条件を付加した鍵の授受（譲渡、貸与）による排他制御方式を用いることにより、特に多数の計算機が互いに同一のデータを何回も参照したり、時系列的にデータが変化していくようなデータベース応用の場合に以下のような点で効率の良い排他制御方式を提供できる。

【0058】（1）複写のための読み出し鍵を用いた任意の時点でのデータ転送、あるいはマルチキャストデータの受信によりデータベースの複製を生成することで、参照の度にデータ転送を行う必要がなく参照速度や通信回線の負荷の点で効率が良い。

【0059】（2）複製データベース生成・維持のためにマルチキャスト方式を用いることでデータの更新値が通信回線上を1度流れるだけで済むので、通信回線の負荷の点で効率が良い。時系列的に変化するデータについては、各計算機が任意の時点から、他の計算機の負荷を増すことなく、その複製データの所有者となることができる。

【0060】（3）各計算機が複製データベースを所持し、マルチキャスト方式により不特定の計算機に鍵を要求するので、データの所在を管理するディレクトリ・マネージャやデータの排他的な操作を管理するロックマネージャが不要となる。

【図面の簡単な説明】

【図1】本発明の実施形態を示す分散型データベース構成例。

【図2】実施形態における型Aの定義とデータa_iの配布例。

【図3】実施形態におけるデータa_iの読み出し例。

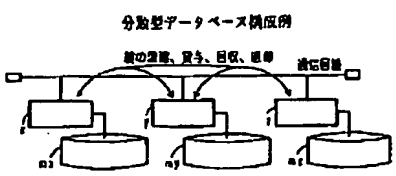
【図4】実施形態におけるデータa_iの書込み例。

【符号の説明】

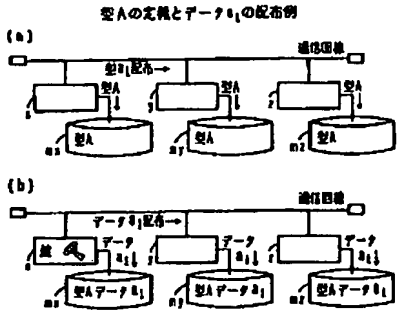
x，y，z…計算機

mx，my，mz…記憶装置

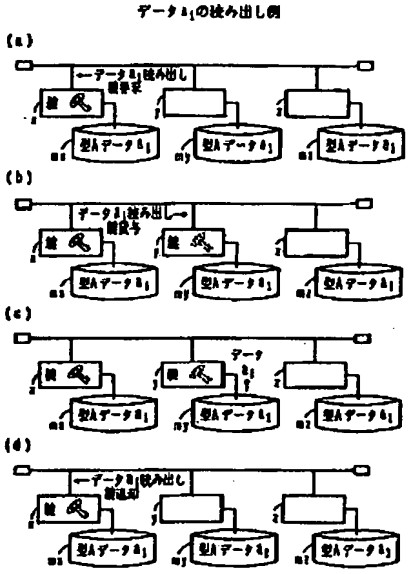
【図1】



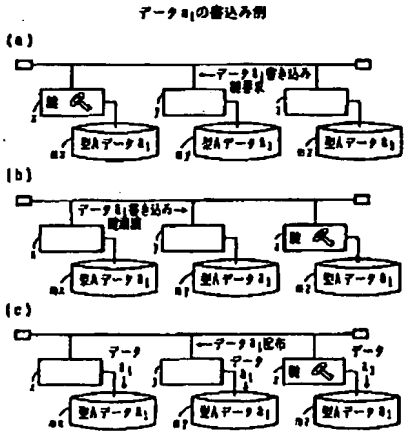
【図2】


型Aの定義とデータa1の配布例
(a)
通信回線
型A
(b)
通信回線

【図3】


データa1の読み出し例
(a)
(b)
(c)
(d)

【図4】


データa1の書込み例
(a)
(b)
(c)